Title

Functional organization of human sensorimotor cortex for speech articulation.

Journal

Nature, 495(7441)

ISSN

0028-0836

Authors

Bouchard, Kristofer E
Mesgarani, Nima
Johnson, Keith
et al.

Publication Date

2013-03-01

DOI

10.1038/nature11911

Peer reviewed

HHS Public Access

Author manuscript

Nature. Author manuscript; available in PMC 2013 September 21.

Published in final edited form as:

Nature. 2013 March 21; 495(7441): 327–332. doi:10.1038/nature11911.

Functional Organization of Human Sensorimotor Cortex for Speech Articulation

Kristofer E. Bouchard[1,2], **Nima Mesgarani**[1,2], **Keith Johnson**[3], and **Edward F. Chang**[1,2,4]

[1]Departments of Neurological Surgery and Physiology, University of California, San Francisco

[2]Center for Integrative Neuroscience, University of California, San Francisco

[3]Department of Linguistics, University of California, Berkeley

[4]UCSF Epilepsy Center, University of California, San Francisco

Abstract

Speaking is one of the most complex actions we perform, yet nearly all of us learn to do it effortlessly. Production of fluent speech requires the precise, coordinated movement of multiple articulators (e.g., lips, jaw, tongue, larynx) over rapid time scales. Here, we used high-resolution, multi-electrode cortical recordings during the production of consonant-vowel syllables to determine the organization of speech sensorimotor cortex in humans. We found speech articulator representations that were somatotopically arranged on ventral pre- and post-central gyri and partially overlapping at individual electrodes. These representations were temporally coordinated as sequences during syllable production. Spatial patterns of cortical activity revealed an emergent, population-level representation, which was organized by phonetic features. Over tens of milliseconds, the spatial patterns transitioned between distinct representations for different consonants and vowels. These results reveal the dynamic organization of speech sensorimotor cortex during the generation of multi-articulator movements underlying our ability to speak.

Speech communication critically depends on the capacity to produce the large set of sounds that compose a given language[1,2]. The wide range of spoken sounds results from highly flexible configurations of the vocal tract, which filters sound produced at the larynx via precisely coordinated movements of the lips, jaw and tongue [3–5]. Each articulator has extensive degrees of freedom, allowing a large number of different realizations for speech movements. How humans exert such exquisite control in the setting of highly variable movement possibilities is a central unanswered question[1,6,7].

Correspondence and requests for materials should be addressed to E.F.C. (ChangEd@neurosurg.ucsf.edu).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Author Contributions

E.F.C. conceived and collected the data for this project. K.E.B. designed and implemented the analysis with assistance from E.F.C. N.M. assisted with preliminary analysis. K.E.B. and E.F.C. wrote the manuscript. K.J. provided phonetic consultation on experimental design and interpretation of results. E.F.C. supervised the project.

The cortical control of articulation is primarily mediated by the ventral half of the lateral sensorimotor (Rolandic) cortex (ventral sensorimotor cortex, vSMC) [11–13,17–19], which provides corticobulbar projections to and afferent innervation from the face and vocal tract (Fig. 1a, b)[8,9]. The U-shaped vSMC is composed of the pre- and post-central gyri (Brodmann areas 1, 2, 3, 6b), and the gyral area directly ventral to the termination of the central sulcus called the guenon (Brodmann area 43) (Fig. 1a, b) [8,10–13]. Using electrical stimulation, Foerster and Penfield described the somatotopic organization of face and mouth representations in human vSMC[14,15]. However, focal stimulation could not evoke meaningful utterances[14,16], implying that speech is not stored in discrete cortical areas. Instead, the production of phonemes and syllables is thought to arise from a coordinated motor pattern involving multiple articulator representations[1,3,4,12].

To understand the functional organization of vSMC in articulatory sensorimotor control, we recorded neural activity directly from the cortical surface in three human subjects implanted with high-density multi-electrode arrays as part of their work-up for epilepsy surgery[17] (Fig. 1a). Intracranial cortical recordings were synchronized with microphone recordings as subjects read aloud consonant-vowel (CV) syllables (19 consonants followed by either /a/, /u/, or /i/, Supplementary Figure 1) commonly used in American English. This task was designed to sample across a range of phonetic features, including different constriction locations (place of articulation), and different constriction degree/shape (manner of articulation) for a given articulator [18–20].

vSMC Physiology During Syllable Production

We aligned cortical recordings to acoustic onsets of consonant-to-vowel transitions (t = 0) to provide a common reference point across CVs (Fig. 1c–e). We focused on the high-gamma frequency component of local field potentials (85–175 Hz)[17,21,22], which correlates well with multi-unit firing rates[23]. For each electrode, we normalized the time-varying high-gamma amplitude to baseline statistics by transforming to z-scores.

During syllable articulation, approximately 30 active vSMC electrode sites were identified per subject (~1200 mm^2, change in z-score > 2 for any syllable). Cortical activity from selected electrodes distributed along the vSMC dorso-ventral axis is shown for /ba/, /da/, and /ga/ (Fig. 1c–e). The plosive consonants (/b/, /d/, /g/) are produced by transient occlusion of the vocal tract by the lips, front tongue, and back tongue, respectively, whereas the vowel /a/ is produced by a low, back tongue position during phonation. Dorsally located electrodes (e.g. e124, e108; black) were active during production of /b/, which requires transient closure of the lips. In contrast, mid-positioned electrodes (e.g. e129, e133, e105; grey) were active during production of /d/, which requires forward tongue protrusion against the alveolar ridge. Amore ventral electrode (e.g. e104; red) was most active during production of /g/, which requires a posterior-oriented tongue elevation towards the soft palate. Other electrodes appear to be active during the vowel phase for /a/ (e.g. e154, e136, e119).

Cortical activity at different electrode subsets was superimposed to visualize spatiotemporal patterns across other phonetic contrasts. Consonants produced with different constriction

locations of the tongue tip, (e.g. /θ/ [dental], /s/ [alveolar], and /ʃ/ [post-alveolar]), showed specificity across different electrodes in central vSMC (Fig. 1f), though were not as categorical as those shown for articulators shown in Fig. 1c–e. Consonants with similar tongue constriction locations, but different constriction degree/shape, were generated by overlapping electrode sets exhibiting different relative activity magnitudes (Fig. 1g, /l/ [lateral] vs. /n/ [nasal stop] vs. /d/ [oral stop]). Syllables with the same consonant followed by different vowels (Fig. 1h, /ja/, /ji/, /ju/) revealed similar activity patterns preceding the CV transition. During vowel phonation, a dorsal electrode is clearly active during /u/ (black arrow), but not /i/ or /a/, whereas another electrode in the middle of vSMC had prolonged activity during /i/ and /u/ vowels compared to /a/ (red arrow). These contrastive examples illustrate that important phonetic properties can be observed qualitatively from the rich repertoire of vSMC spatiotemporal patterns.

Spatial Representation of Articulators

To determine the spatial organization of speech articulator representations, we examined how cortical activity at each electrode depended upon the movement of a given articulator (using a general linear model). We assigned binary variables to four articulatory organs (lips, tongue, larynx, and jaw) used in producing the consonant component of each CV (Supplementary Figure 1). In Figure 2a, we plot spatial distribution of optimal weightings for these articulators (averaged over time and subjects), as a function of dorsal-ventral distance from the Sylvian fissure and anterior-posterior distance from the central sulcus. We found representations for each articulator distributed across vSMC (Fig. 2a). For example, the lip representation was localized to the dorsal aspect of vSMC, whereas the tongue representation was more distributed across the ventral aspect.

To determine topographic organization of articulators across subjects, we extracted the greatest 10% of weightings from individual articulator distributions (Fig. 2a) and used a clustering algorithm (k-nearest neighbor) to classify the surrounding cortex (Fig. 2b). We found an overall somatotopic dorsal-ventral arrangement of articulator representations laid out in the following sequence: larynx (X), lips (L), jaw (J), tongue (T), and larynx (X) (Fig. 2a, b, see Supplementary Figures 2–5). An analysis of the fractional representation of all articulators at single electrodes revealed clear preferred tuning for individual articulators at single electrodes and also demonstrated that single electrodes had functional representations of multiple articulators (Supplementary Figure 6).

Timing of Articulator Representations

Because the time-course of articulator movements is on the scale of tens of milliseconds, previous approaches have been unable to resolve temporal properties associated with individual articulator representations. We examined the timing of correlations between cortical activity and specific consonant articulators (using partial correlation analysis), and included two vowel articulatory features (back tongue & high tongue; Supplementary Figure 1).

In figure 3a, we plot time courses of correlations for electrodes with highest values, sorted by onset latency. We found that jaw, high tongue, and back tongue had very consistent

timing across electrodes. Similar results were found for tongue, lips, and larynx, but with more variable latencies. Timing relationships between articulator representations were staggered, reflecting a temporal organization during syllable production: lip and tongue correlations began well before sound onset (Fig. 3a, c, d); jaw and larynx correlations were aligned to the CV transition (Fig. 3a, c, d); high tongue and back tongue features showed high temporal specificity for the vowel phase, peaking near the acoustic mid-point of the vowels (~250 ms, Fig. 3b–d). This sequence of articulator correlations was consistent across subjects (Fig. 3d, $P < 10^{-10}$, ANOVA, $F = 40$, $df = 5$, $n = 211$ electrodes from 3 subjects) and accords with timing of articulator movement shown in speech kinematic studies[3,5,18,24]. We found no significant onset latency differences in those areas immediately anterior and posterior to the central sulcus (±10 mm), or across the geunon (Supplementary Figure 7). This is consistent with mixed sensory and motor orofacial responses throughout vSMC, which are also seen in stimulation experiments[14,25]

Phonetic Organization of Spatial Patterns

We hypothesized that the coordination of multiple articulators required for speech production would manifest as spatial patterns of cortical activity. Here, we refer to this population-derived pattern as the phonetic representation. To determine its organizational properties, we used principal components analysis to transform the observed cortical activity patterns into a 'cortical state-space' (9 spatial PCs for all subjects, ~60% of variance explained, Supplementary Figure 8–9)[26–30]. K-means clustering during the consonant phase (−25 ms prior to the release of the consonant) showed that the cortical state-space was organized according to the major oral articulators (quantified by Silhouette analysis): labial (lips), coronal (front) tongue, and dorsal (back) tongue (Fig. 4a, Supplementary Figure 10). During the vowel phase, we found clear separation of /a/, /i/, and /u/ vowel states (Fig. 4b). Similar clustering of consonants and vowels was found across subjects ($P < 10^{-10}$ for clustering of both consonants and vowels, Supplementary Figure 11).

Theories of speech motor control and phonology have speculated a hierarchical organization of phoneme representations given anatomical and functional dependencies of the vocal tract articulators during speech production[1,3,4,18,19,31]. To evaluate such organization in vSMC, we applied hierarchical clustering to the cortical state-space (Fig. 4c–d). For consonants, this analysis confirmed that the primary tier of organization was defined by the major oral articulator features: dorsal, labial, or coronal (Fig. 4c). These major articulators were superordinate to the constriction location within each articulator. For example, the labial cluster could be subdivided into bi-labial and labio-dental. Only at the lowest level did we observe suggestions of organization according to constriction degree/shape, such as the sorting of nasal (/n/ syllables), oral stops (/d/, /t/), and lateral approximants (/l/). Analogously, during the vowel period, a primary distinction was based upon the presence/absence of lip-rounding (/u/ vs. /a/ & /i/) and secondary distinction based on tongue posture (height and front/back position)(Fig. 4d). Therefore, the major oral articulator features that organize consonant representations are similar to those for vowels.

Across early time points (−375:120 ms), we found that consonant features describing constriction location yielded a significantly greater correlation with the cortical state-space

than constriction degree, which in turn was significantly more correlated than the upcoming vowel ($P < 10^{-10}$, Wilcoxon signed-rank test (WSRT), n = 297 from 3 subjects, see Supplementary Figure 12 for phonetic feature sets). This analysis demonstrates that constriction location accounts for more structure of spatial activity patterns than does constriction degree/shape. Analogously, across later time points (125:620 ms), we found that vowel features provided the greatest correlation (vowel configuration vs. others, $P < 10^{-10}$, WSRT, n = 297 from 3 subjects).

Dynamics of Phonetic Representations

The dynamics of neural populations have provided insight into the structure and function of many neural circuits [6,26,27,29,32,33]. To determine the dynamics of phonetic representations, we investigated how state-space trajectories for CVs entered and departed target regions for phonetic clusters. Trajectories of individual CV syllables were visualized by plotting their locations in the first two PC dimensions versus time (Fig. 5a, b, PC1 & PC2; for Subject 1).

We first examined how trajectories of different consonants transitioned to a single vowel /u/ (Figure 5a). The cortical state-space was initially unstructured, and then individual trajectories converged within phonetic clusters (e.g. labial, front tongue, dorsal tongue, and sibilant), while simultaneously cluster trajectories diverged from one another. These convergent and divergent dynamics gradually increased the separability of different phonetic clusters. Later, as each consonant transitioned to /u/, trajectories converged to a compact target region for the vowel. Finally, trajectories diverged randomly, presumably as articulators returned to neutral position. Analogous dynamics were observed during the production of a single consonant cluster (e.g. labials) transitioning to different vowels (/a/, /i/, and /u/)(Fig. 5b).

We quantified the internal dynamical properties of the cortical state-space by calculating cluster separability, which measures the mean difference of between-cluster and within-cluster distances. The time course of cluster separability, averaged across subjects and CVs, is plotted in Figure 5c: separability peaked ~200 ms before the CV transition for consonants (onset, ~−300 ms), and at +250 ms for vowels (onset, ~50 ms). We further examined the dynamics of correlations between the structure of the cortical state-space and phonetic features (averaged across subjects), which is plotted in Figure 5d. Across subjects, we found that cluster separability and the correlation between cortical state-space organization and phonetic features were tightly linked for both consonants and vowels in a time-dependent fashion (R^2 range = [0.42–0.98], $P < 10^{-10}$ for all). This demonstrates that the dynamics of clustering in the cortical state-space is strongly coupled to the degree to which the cortical state reflects the phonetic structure of the vocalization.

The dynamic structure of the cortical state-space during production of all CV syllables is summarized in Figure 5e. In this visualization, the center of each colored tube is located at the centroid of the corresponding phonetic cluster. Tube diameter corresponds to cluster density and color saturation represents the correlation between the structure of the cortical state-space and phonetic features. This visualization highlights that, as the cortical state comes to reflect phonetic structure (coloring), different phonetic clusters diverged from one

another, while the trajectories within clusters converged. Furthermore, we observed correlates of the earlier articulatory specification for sibilants (red, e.g. /sh/, /z/, /s/). Additionally, with all CVs on the same axes, we observed that consonants occupy a substantially distinct region of cortical state-space compared to vowels, despite sharing the same articulators. The distribution of distances comparing consonant and vowel representations was significantly greater than the consonant-consonant comparison or vowel-vowel comparison ($P < 10^{-10}$ for all comparisons, WSRT, n = 4623 for all, Supplementary Figure 11). Finally, the consonant-to-vowel sequence reveals a periodic structure, which is sub-specified for consonant and vowel features.

Discussion

Our broad-coverage, high-resolution, direct cortical recordings allowed us to examine the spatial and temporal profiles of speech articulator representations in human vSMC. Cortical representations were somatotopically organized, and punctuated by sites tuned for a preferred articulator and co-modulated by other articulators. The dorsal-ventral layout of articulators recapitulates the rostral-to-caudal layout of the vocal tract. However, we found an additional laryngeal representation located at the dorsal-most end of vSMC [11,13,34,35]. This dorsal laryngeal representation appears to be absent in non-human primates[8,36,37], suggesting a unique feature of vSMC for the specialized control of speech. Pre- and post-central gyrus neural activity occurred before vocalization, which may reflect the integration of motor commands with proprioceptive information for rapid feedback control during speaking[12,38–43].

Just as focal stimulation is insufficient to evoke speech sounds, it is not any single articulator representation, but rather the coordination of multiple articulator representations across the vSMC network that generates speech. Analysis of spatial patterns of activity revealed an emergent hierarchy of network states, which organized phonemes by articulatory features. This functional hierarchy of network states contrasts with the anatomical hierarchy often considered in motor control[44]. The cortical state-space organization likely reflects the coordinative patterns of articulatory motions during speech, and is strikingly similar to a theorized cross-linguistic hierarchy of phonetic features (“feature geometry”)[3,19,20,31,45]. In particular, the findings support gestural theories of speech control[3] over alternative acoustic (hierarchy primarily organized by constriction degree)[20] or vocal tract geometry theories (no hierarchy of constriction location and degree)[19].

The vSMC population exhibited convergent and divergent dynamics during production of different phonetic features. The dynamics of individual phonemes were superimposed on a slower oscillation characterizing the transition between consonants and vowels, which occupied distinct regions of the cortical state-space. Although trajectories could originate or terminate in different regions, they consistently passed through the same (target) region of the state-space for shared phonetic features[46]. Large state-space distances between consonant and vowel representations may explain why it is more common to substitute consonants with one another, and same for vowels, but very rarely across categories in speech errors (i.e. slips of the tongue’)[47].

We have demonstrated that a relatively small set of articulator representations can flexibly combine to create the large variety of speech sounds in American English. The major organizational features found here define many phonologies throughout the world[31]. Consequently, these cortical organizational principles are likely to be conserved, with further specification for unique articulatory properties across different languages.

Methods

The experimental protocol was approved by the Human Research Protection Program at the University of California, San Francisco.

Subjects and Experimental Task

Three native English speaking human subjects underwent chronic implantation of a high-density, subdural electrocortigraphic (ECoG) array over the left hemisphere (two subjects) or right hemisphere (one subject) as part of their clinical treatment of epilepsy (see Supplementary Table 1 for clinical details)[48]. Subjects gave their written informed consent before the day of surgery. All subjects had self-reported normal hearing and underwent neuropsychological language testing (including the Boston Naming and verbal fluency tests) and were found to be normal. Each subject read aloud consonant-vowel syllables (CVs) composed of 18–19 consonants (19 consonants for two subjects, 18 consonants for one subject), followed by one of three vowels. Each CV was produced between 15 and 100 times. Microphone recordings were synchronized with the multi-channel ECoG data.

Data acquisition and Signal Processing

Cortical local field potentials (LFP) were recorded with ECoG arrays and a multi-channel amplifier optically connected to a digital signal processor (Tucker-Davis Technologies [TDT], Alachua, FL). The spoken syllables were recorded with a microphone, digitally amplified, and recorded inline with the ECoG data. ECoG signals were acquired at 3052 Hz.

The time series from each channel was visually and quantitatively inspected for artifacts or excessive noise (typically 60 Hz line noise). These channels were excluded from all subsequent analysis and the raw recorded ECoG signal of the remaining channels were then common average referenced and used for spectro-temporal analysis. For each (useable) channel, the time-varying analytic amplitude was extracted from eight bandpass filters (Gaussian filters, logarithmically increasing center frequencies (85–175 Hz) and semi-logarithmically increasing band-widths) with the Hilbert transform. The high-gamma (high-γ) power was then calculated by averaging the analytic amplitude across these eight bands, and then this signal was down-sampled to 200 Hz. High-γ power was z-scored relative to the mean and standard deviation of baseline data for each channel. Throughout the Methods, when we speak of High-γ power refers to this z-scored measure, denoted below as Hγ.

Acoustic Analysis

The recorded speech signal was transcribed off-line using WaveSurfer (http://www.speech.kth.se/wavesurfer/). The onset of the consonant-to-vowel transition (C->V) was used as the common temporal reference point for all subsequent analysis (see

Supplemental Figure 1). This was chosen because it permits alignment across all of the syllables and allows for a consistent discrimination of the consonantal and vocalic components. Post-hoc analysis of acoustic timing revealed the onset of the consonant-to-vowel transition to be highly reproducible across multiple renditions of the same syllable. As such, alignment at C->V results in relatively small amounts of inter-syllable jitter in estimated times of acoustic onset, offset and peak power.

For temporal analysis of the CV acoustic structure, each individual vocalization was first converted to a cochlear spectrogram by passing the sound-pressure waveform through a filter bank emulating the cochlear transfer function[49]. As the current analysis of cortical data leverages the cross-syllabic variability in (average) Hγ (see below), we reduced the dataset of produced vocalizations to a single exemplar for each CV syllable. For each unique CV syllable, the cochlear spectrograms associated with each utterance of that CV ($S_i(t,f)$) were analyzed to find a single proto-typical example (Pspct), defined as the syllable that had the minimum spectral-temporal difference from every other syllable of that kind:

$$Pspct = min_{S_i}\left(\sum_j \sum_{t,f} (S_j(t,f) - S_i(t,f))^2\right) \quad (1)$$

The onset, peak, and offset of acoustic power were extracted for each syllable prototype using a thresholding procedure.

Articulator state matrix and Phonetic feature matrix

To describe the engagement of the articulators in the production of different CV syllables, we drew from standard descriptions of the individual consonant and vowel sounds in the International Phonetic Alphabet (IPA)[50]. Each CV syllable was associated with a binary vector describing the engagement of the speech articulators utilized to produce the CV. For the linear analysis presented in Figures 2 and 3, the articulator state vector (B_i) for each CV syllable s_i was defined by six binary variables describing the four main articulator organs (Lips, Tongue, Larynx, Jaw) for consonant production and two vocalic tongue configurations (High Tongue, Back Tongue) (Supplemental Figure 1). Although more detailed descriptions are possible (e.g. alveolar-dental), the linear methods utilized for these analyses necessitate that the articulator variables be linearly independent (no feature can be completely described as a linear combination of the others), though the features may have correlations. An expanded phonetic feature matrix (nine consonant constriction location variables, four tongue configuration variables, and six consonant constriction degree/shape variables, again derived from IPA, Supplemental Figure 9), was used in the non-parametric analysis of the cortical state-space (Figures 4 and 5).

Analysis of Articulator Representations

Spatial Organization Derived from a General Linear Model

To examine the spatial organization with which Hγ was modulated by the engagement of the articulators, we determined how the activity of each electrode varied with consonant articulator variables using a general linear model. Here, at each moment in time (t), the

general linear model described the Hγ of each electrode as an optimally weighted sum of the articulators engaged during speech production. Hγ(t) recorded on each electrode (e_i), during the production of syllable s_j, $H\gamma_{ij}(t)$, was modeled as a linear weighted sum of the binary vector associated with the consonant component of s_j,(B^c_j):

$$H\gamma_{i,j}(t)=\beta_i(t)\cdot B^C_j+\beta_{i0}(t) \quad (2)$$

The coefficient vector $\beta_i(t)$ that resulted in the least-mean square difference between the levels of activity predicted by this model and the observed Hγ(t) across all syllables was found by linear regression. For each electrode e_i at time t, the associated 1×4 slope vector ($\beta_i(t)$) quantifies the degree to which the engagement of a given articulator modulated the cross-syllable variability in Hγ(t) at that electrode. Coefficients of determination (R^2) were calculated from the residuals of this regression. In the current context, R^2 can be interpreted as the amount of cross-syllabic variability in Hγ that can be explained by the optimally weighted linear combination of articulatory state variables.

The spatial organization of the speech articulators was examined using the assigned weight vectors ($\beta_i(t)$) from the GLM described above. First, the fit of the GLM at each electrode e_i was determined of interest if, on average, the associated p-value was less than 0.05 for any one of the four consonant articulator time windows (T_A) determined from the partial-correlation analysis below. We defined this time window to be the average onset-to-offset time of significant partial correlations for each individual articulator in each subject (see Partial Correlation Analysis). This method identifies electrodes whose activity is well predicted by the GLM for any of the individual articulators, as well as combinations thereof, for extended periods of time. As these time windows extend for many points this is a rather stringent criterion in comparison to a peak-finding method or single significant-crossings. In practice, the minimum (across time) p-values associated with the vast majority of these electrodes are several orders of magnitude less then 0.05. For the electrodes gauged to be significant in each subject, we averaged the weights for each articulator (A) in that articulators time window (T_A). Thus, each electrode of interest (e_i) is assigned four values, with each value corresponding to the weighting for that articulator (A), averaged across that articulator's time window (T_A):

$$W^A_i=\frac{1}{|T_A|}\sum_{t\in T_A}\beta_i(t) \quad (3)$$

For the analysis of representational overlap at individual electrodes (Figure 2c), each electrode was classified according to the dominant articulator weight in a winner-take-all manner[48]. The fractional articulator weighting was calculated based off of the positive weights at each electrode, and is plotted as average percent of summed positive weights.

For spatial analysis, the data for each subject was smoothed using a 2mm uniform circular kernel. The maps presented and analyzed in Supplementary Figure 2–3 correspond to these average weights for the Lips, Tongue, Larynx, and Jaw. The maps presented and analyzed in Figure 2 correspond to these average weights for each articulator averaged across subjects.

The spatial organization of vSMC is described by plotting the results of the GLM for an individual on the cortex of that individual. We used a Cartesian plane defined by the Anterior-Posterior distance from the Central Sulcus (ordinate) and the Dorsal-Ventral distance from the Sylvian Fissure (azimuth). This provides a consistent reference frame to describe the spatial organization of each subject's cortex and to combine data across subjects while preserving the individual differences.

Somatotopic map and k -Nearest Neighbors Algorithm

To construct the summary somatotopic map of Figure 2b, we first extracted the spatial location of the top 10% of weights for each articulator (averaged across subjects, data in Figure 2a). We then used a k-nearest neighbor algorithm to classify the surrounding cortical tissue based on the nearest k = 4 neighbors within a spatial extent of 3 mm of each spatial location; if no data points were present within 3 mm, the location is unclassified (white). Locations where no clear majority (> 50%) of the nearest-neighbors belonged to a single articulator were classified as mixed (gold). These values were chosen to convey, in summary form, the visual impression of the individual articulator maps, and to 'fill-in' spatial gaps in our recordings. The summary map changed smoothly and as expected with changes in threshold of individual articulator maps, k (number of neighbors), spatial extent, and minimum number of points. Results are qualitatively insensitive to the details of this analysis, including the choice of 10% as a threshold, as changes in the clustering algorithm could be made to accommodate subtle differences in data inclusion. For visual comparison, we display the somatotopic maps derived from the same algorithm derived from the top 5%, top 10% and top 15% of weights in Supplementary Figure 4.

Partial Correlation Analysis

To quantify the temporal structure with which single-electrode Hγ was correlated with the engagement of a single articulator, we used partial correlation analysis. Partial correlation analysis is a standard statistical tool that quantifies the degree of association between two random variables (here, $H\gamma(t)$ and the engagement of a given articulator, A_i), while removing the effect of a set of other random variables (here, the other articulators, A_j, $j \neq i$). For a given electrode, the partial correlation coefficient between $H\gamma(t)$ across syllables at time t and articulator A_i ($\rho(H\gamma(t),A_i)$) is calculated as the correlation coefficient between the residuals $r(H\gamma(t),A_j)$, $j \neq i$, resulting from de-correlating the $H\gamma(t)$ and every other articulator A_j, $i \neq j$, and the residuals $r(A_i,A_j)$, $i \neq j$, resulting from de-correlating the articulators from one another:

$$\rho(H\gamma(t), A_i) = \frac{cov(r(H\gamma(t), A_j), r(A_i, A_j))}{\sigma_1 \cdot \sigma_2}, i \neq j \quad (4)$$

Where σ_1 and σ_2 are the standard deviations of $r(H\gamma(t),A_j)$ and $r(A_i,A_j)$ respectively. In the current context, the partial correlation coefficients quantify the degree to which the cross-syllabic variability in Hγ at a given moment in time was uniquely associated with the engagement of a given articulator during speech production. For each articulator, we analyzed those electrodes whose peak partial correlation coefficient (ρ) exceeded the mean ± 2.5 σ of ρ values across electrodes and time (> mean($\rho(e_i,t)$)+ 2.5σ($\rho(e_i,t)$)). In the text, we

focus on the positive correlations (which we denote as R), because there were typically a larger number of positive values (mean $\rho > 0$), the temporal profiles are grossly similar for negative values, and for expositional simplicity. Results did not qualitatively change with changes in this threshold of ~ ± 0.2 σ. We extracted the onset, offset, and peak times for each articulator for each electrode that crossed this threshold. The data presented in Figure 3d are the mean ± s.e. of these timing variables across electrodes pooled across subjects. The average onset and offset for each of the four consonant articulators (A [Lips, Tongue, Jaw, Larynx]) in each subject was used to define the articulator time window (T_A) used in the spatial analysis described above.

Cortical State-Space and State-Space Analysis

PCA and Cortical-State Space

Principal components analysis (PCA) was performed on the set of all vSMC electrodes for dimensionality reduction and orthogonalization. PCA was performed on the n x m*t covariance matrix **Z** with rows corresponding to channels (of which there are n) and columns corresponding to concatenated $H\gamma(t)$ (length t) for each CV (of which there are m). Each electrode's time series was z-scored across syllables to normalize response variability across electrodes. The singular-value decomposition of **Z** was used to find the eigenvector matrix **M** and associated eigenvalues λ. The PCs derived in this way serve as a spatial filter of the electrodes, with each electrode e_j receiving a weighting in PC_i equal to $\mathbf{M}_{ij}$, where **M** is the matrix of eigenvectors. Across subjects, we observed that the eigenvalues (λ) exhibited a fast decay with a sharp inflection point at the ninth eigenvalue, followed by a much slower decay thereafter (Supplemental Figure 6). We therefore used the first nine eigenvectors (PC's) as the cortical state-space for each subject.

The cortical state-space representation of syllable s_k at time t, $\mathbf{K}(s_k,t)$, is defined as the projection of the vector of cortical activity associated with s_k at time t, $H\gamma_k(t)$, onto **M**:

$$\mathbf{K}(s_k, t) = \mathbf{M} \cdot H\gamma_k(t) \quad (5)$$

We calculated the contribution of articulators to the cortical state-space (PCw_{ij}) by projecting each electrodes weight vector (β_j) derived from the GLM model above into the first three dimensions of the cortical state-space (i=1:3):

$$PCw_{ij} = \mathbf{M}_{ij} \cdot \beta_j \quad (6)$$

Here, PCw_{ij} is a four-element vector of the projected articulator weights for electrode e_j into PC_i. In Supplemental Figure 8, we plot $\log_{10}$ of the absolute value of PCw_{ij} across electrodes, which describes the distribution of magnitudes of the representations associated with the four articulators in a given PC.

Clustering Analysis

k-means and hierarchical clustering were performed on the cortical state-space representations of syllables, $\mathbf{K}(s_k,t)$, based on the pair-wise Euclidean distances calculated

between CV syllable representations. Agglomerative hierarchical clustering used Ward's Method. All analyses of the detailed binary phonetic feature matrix were performed using both Hamming and Euclidean distances; results did not change qualitatively or statistically between metrics. We utilized silhouette analysis to validate the claim that there were three clusters at the consonant time. The silhouette of a cluster is a measure of how close (on average) the members of that cluster are to each other, relative to the next nearest cluster. For a particular data set, the average silhouette for a given number of clusters describes the parsimony of the number of clusters in the data. Hence, examining the silhouette across different number of clusters gives a quantitative way to determine the most parsimonious number of clusters[51]. Higher values correspond to more parsimonious clustering. On average across subjects, this analysis validated the claim that three clusters (Average Silhouette = 0.47) was a more parsimonious clustering scheme than either two (Average Silhouette = 0.45) or four clusters (Average Silhouette = 0.43).

Correlation of cortical state-space structure with phonetic structure

At each moment in time, we wanted to quantify the similarity of the structure of cortical state-space representations of phonemes and the structure predicted by different phonetic feature sets. To this end, we measured the linear correlation coefficient between vectors of unique pair-wise Euclidean distances between phonemes calculated in the cortical state-space (DC(t)) and in the phonetic feature matrix (DP):

$$R(t)=\frac{cov\,(DC(t),DP)}{\sigma_{DC(t)}\cdot\sigma_{DP}} \quad (7)$$

As described above, the phonetic feature matrix was composed of three distinct phonetic feature sets, (consonant constriction location, consonant constriction degree/shape, vowel configuration). Distances were calculated independently in these three sub-sets and correlated with DC. Standard error measures of the correlation coefficients were calculated using a bootstrap procedure (1000 iterations).

Cluster Separability

Cluster separability is defined at any moment in time as the difference between the average of cross-cluster distances and the average of within cluster distances. This quantifies the average difference of the distance between syllables in different clusters and the tightness of a given cluster. We quantified the variability in cluster separability estimation through a 1000 iteration bootstrap procedure of the syllables used to calculate the metric.

Cluster Density

We quantified the average cluster density by calculating the average inverse of all unique pair-wise distances between CVs in a given cortical state-space cluster. It is a proper density because the number of elements in a cluster does not change with time.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Angela Ren for technical help with data collection and pre-processing, and Miranda Babiak for audio transcription. J. Houde, C. Niziolek, S. Lisberger, K. Chaisanguanthum, C. Cheung, and I. Garner for helpful comments on the manuscript. E.F.C. was funded by National Institutes of Health grants R00-NS065120, DP2-OD00862, R01-DC012379, and the Ester A. and Joseph Klingenstein Foundation.

References

1. Levelt, WJM. Speaking: From Intention to Articulation. MIT Press; 1993.
2. Ladefoged, P.; Johnson, K. A Course in Phonetics. Wadsworth Publishing; 2010.
3. Browman CP, Goldstein L. Articulatory Gestures as Phonological Units. Haskins Laboratories Status Report on Speech Research. 1989; 99:69–101.
4. Fowler, CA.; Rubin, PE.; Remez, RE.; Turvey, MT. Language Production. In: Butterworth, B., editor. Speech and Talk. Vol. 1. Academic Press; 1980. p. 373-420.
5. Gracco VL, Lofqvist A. Speech motor coordination and control: evidence from lip, jaw, and laryngeal movements. J Neurosci. 1994; 14:6585–6597. [PubMed: 7965062]
6. Schoner G, Kelso JA. Dynamic pattern generation in behavioral and neural systems. Science. 1988; 239:1513–1520. [PubMed: 3281253]
7. Franklin DW, Wolpert DM. Computational mechanisms of sensorimotor control. Neuron. 2011; 72:425–442. [PubMed: 22078503]
8. Jurgens U. Neural pathways underlying vocal control. Neurosci Biobehav Rev. 2002; 26:235–258. [PubMed: 11856561]
9. Kuypers HG. Corticobular connexions to the pons and lower brain-stem in man: an anatomical study. Brain. 1958; 81:364–388. [PubMed: 13596471]
10. Brodmann, K. Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues. Smith-Gordon; 1909/1994.
11. Brown S, et al. The somatotopy of speech: phonation and articulation in the human motor cortex. Brain Cogn. 2009; 70:31–41. [PubMed: 19162389]
12. Guenther FH, Ghosh SS, Tourville JA. Neural modeling and imaging of the cortical interactions underlying syllable production. Brain Lang. 2006; 96:280–301. [PubMed: 16040108]
13. Schulz GM, Varga M, Jeffires K, Ludlow CL, Braun AR. Functional neuroanatomy of human vocalization: an H215O PET study. Cereb Cortex. 2005; 15:1835–1847. [PubMed: 15746003]
14. Penfield W, Boldrey E. Somatic Motor and Sensory Representation in The Cerebral Cortex of Man Studied by Electrical Stimulation. Brain. 1937; 60:389–443.
15. Foerster O. The cerebral cortex in man. Lancet. 1931; 221:309–312.
16. Penfield, W.; Roberts, R. Speech and Brain: Mechanisms. Princeton; 1959.
17. Mesgarani N, Chang EF. Selective cortical representation of attended speaker in multi-talker speech perception. Nature. 2012; 485:233–236. [PubMed: 22522927]
18. Saltzman EL, Munhall KG. A Dynamical Approach to Gestural Patterning in Speech Production. Ecological Psychology. 1989; 1:333–382.
19. Clements, GN.; Hume, E. Handbook of Phonological Theory. Goldsmith, J., editor. Oxford: Basil Blackwell; 1995. p. 245-306.
20. Chomsky, N.; Halle, M. The Sound Pattern of English. The MIT Press; 1991.
21. Crone NE, Miglioretti DL, Gordon B, Lesser RP. Functional mapping of human sensorimotor cortex with electrocorticographic spectral analysis. II. Event-related synchronization in the gamma band. Brain. 1998; 121 (Pt 12):2301–2315. [PubMed: 9874481]
22. Edwards E, et al. Spatiotemporal imaging of cortical activation during verb generation and picture naming. Neuroimage. 2010; 50:291–301. [PubMed: 20026224]
23. Ray S, Maunsell JH. Different origins of gamma rhythm and high-gamma activity in macaque visual cortex. PLoS Biol. 2011; 9:e1000610. [PubMed: 21532743]
24. Kent, RD. The Production of Speech. MacNeilage, PF., editor. Springer-Verlag; 1983.

25. McCarthy G, Allison T, Spencer DD. Localization of the face area of human sensorimotor cortex by intracranial recording of somatosensory evoked potentials. J Neurosurg. 1993; 79:874–884. [PubMed: 8246056]
26. Afshar A, et al. Single-trial neural correlates of arm movement preparation. Neuron. 2011; 71:555–564. [PubMed: 21835350]
27. Mazor O, Laurent G. Transient dynamics versus fixed points in odor representations by locust antennal lobe projection neurons. Neuron. 2005; 48:661–673. [PubMed: 16301181]
28. Sussillo D, Abbott LF. Generating coherent patterns of activity from chaotic neural networks. Neuron. 2009; 63:544–557. [PubMed: 19709635]
29. Ahrens MB, et al. Brain-wide neuronal dynamics during motor adaptation in zebrafish. Nature. 2012; 485:471–477. [PubMed: 22622571]
30. Churchland MM, Cunningham JP, Kaufman MT, Ryu SI, Shenoy KV. Cortical preparatory activity: representation of movement or first cog in a dynamical machine? Neuron. 2010; 68:387–400. [PubMed: 21040842]
31. McCarthy JJ. Feature geometry and dependency: A review. Phonetica. 1988; 45:84–108.
32. Briggman KL, Kristan WB. Multifunctional pattern-generating circuits. Annu Rev Neurosci. 2008; 31:271–294. [PubMed: 18558856]
33. Churchland MM, et al. Neural population dynamics during reaching. Nature. 2012; 487:51–56. [PubMed: 22722855]
34. Brown S, Ngan E, Liotti M. A larynx area in the human motor cortex. Cereb Cortex. 2008; 18:837–845. [PubMed: 17652461]
35. Terumitsu M, Fujii Y, Suzuki K, Kwee IL, Nakada T. Human primary motor cortex shows hemispheric specialization for speech. Neuroreport. 2006; 17:1091–1095. [PubMed: 16837833]
36. Hast MH, Fischer JM, Wetzel AB, Thompson VE. Cortical motor representation of the laryngeal muscles in Macaca mulatta. Brain Res. 1974; 73:229–240. [PubMed: 4208647]
37. Jurgens U. On the elicitability of vocalization from the cortical larynx area. Brain Res. 1974; 81:564–566. [PubMed: 4215545]
38. Pruszynski JA, et al. Primary motor cortex underlies multi-joint integration for fast feedback control. Nature. 2011; 478:387–390. [PubMed: 21964335]
39. Hatsopoulos NG, Suminski AJ. Sensing with the motor cortex. Neuron. 2011; 72:477–487. [PubMed: 22078507]
40. Tremblay S, Shiller DM, Ostry DJ. Somatosensory basis of speech production. Nature. 2003; 423:866–869. [PubMed: 12815431]
41. Matyas F, et al. Motor control by sensory cortex. Science. 2010; 330:1240–1243. [PubMed: 21109671]
42. Rathelot JA, Strick PL. Muscle representation in the macaque motor cortex: an anatomical perspective. Proc Natl Acad Sci U S A. 2006; 103:8257–8262. [PubMed: 16702556]
43. Gracco VL, Abbs JH. Dynamic control of the perioral system during speech: kinematic analyses of autogenic and nonautogenic sensorimotor processes. J Neurophysiol. 1985; 54:418–432. [PubMed: 4031995]
44. Sherrington, CS. The integrative action of the nervous system. Yale University Press; 1911.
45. Jakobson, R.; Fant, G.; Halle, M. Preliminaries to speech analysis: the distinctive features and their correlates. M.I.T. Press; 1969.
46. Keating, PA. The Window model of coarticulation: articulatory evidence. Cambridge University Press; 1990.
47. Dell GS, Juliano C, Govindjee A. Structure and content in language production - a theory of frame constraints in phonological speech errors. Cognitive Science. 1993; 17:149–195.
48. Mesgarani N, Chang EF. Selective cortical representation of attended speaker in multi-talker speech perception. Nature. 2012; 485:233–236. [PubMed: 22522927]
49. Yang X, et al. Auditory representations of acoustic signals. Information Theory, IEEE Transactions on. 1992; 38:824–839.
50. Handbook of the International Phonetic Association. Cambridge University Press; 1999.

51. Rousseeuw PJ. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of computational and applied mathematics. 1987; 20:53–65.

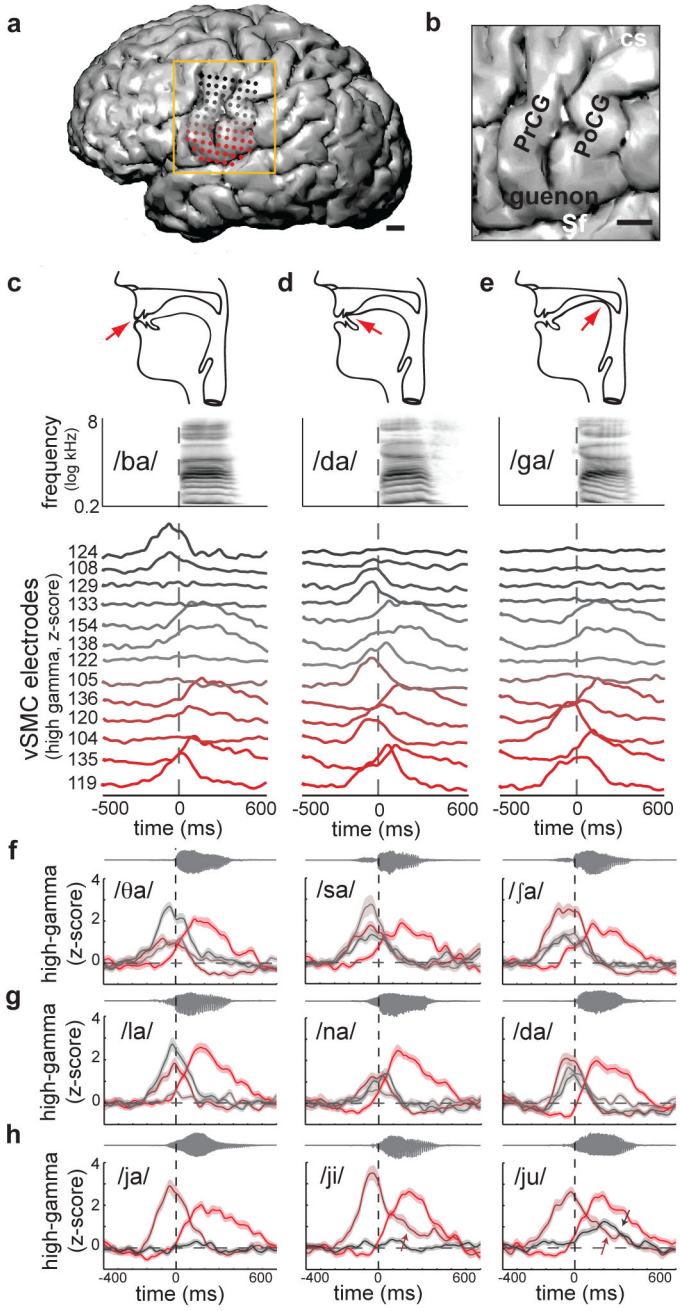

Figure 1. vSMC Physiology During Syllable Production

a, MRI reconstruction of single subject brain with vSMC electrodes (dots), colored by distance from Sylvian fissure. **b,** Expanded view of vSMC anatomy: pre- and post-central gyri (PrCG and PoCG), central sulcus (cs), Sylvian fissure (Sf). Scale bar =1 cm. **c–e** (top), Vocal tract schematics for three consonants (/b/, /d/, /g/), produced by occlusion at the lips, tongue tip, and tongue body, respectively (red arrow). (middle) Spectrograms of spoken consonant-vowel (CV) syllables. (bottom) Average cortical activity from subset of electrodes. Vertical, dashed line is acoustic onset of CV transition. **f–h,** Cortical activity at selected electrodes for different phonetic contrasts (mean ± s.e.). Acoustic waveforms

displayed above. **f,** Fricatives [/θ/(“th” of “thin”), /s/, /ʃ/(“sh” of “shin”)] with different constriction locations. **g,** Front tongue consonants (/l/, /n/, /d/) with different constriction degree/shapes. **h,** Single consonant [/j/ (“y” of “yes”)] with different vowels (/a/, /i/, /u/). Red arrow corresponds to a tongue electrode with prolonged activity for /i/ and /u/ vowels. Black arrow corresponds to active lip electrode for /u/.

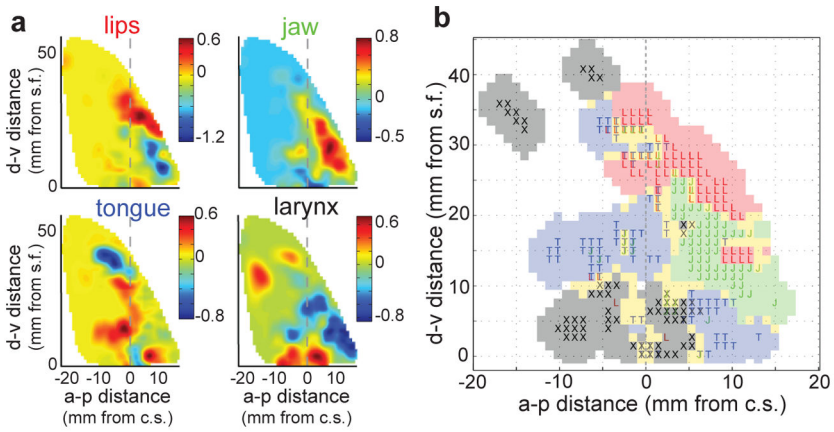

Figure 2. Spatial Representation of Articulators

a, Localization of lips, jaw, tongue, and larynx representations. Average magnitude of articulator weightings (color scale) plotted as a function of anterior-posterior (AP) distance from the central sulcus and dorsal-ventral (DV) distance from the Sylvian fissure (n = 3 subjects). **b,** Functional somatotopic organization of speech articulator representations in vSMC. Lips (L, red); jaw (J, green); tongue (T, blue); larynx (X, black), mixed (Gold). Letters correspond to locations based upon direct measurement-derived regression weights, shaded rectangles correspond to regions classified by k-nearest neighbor.

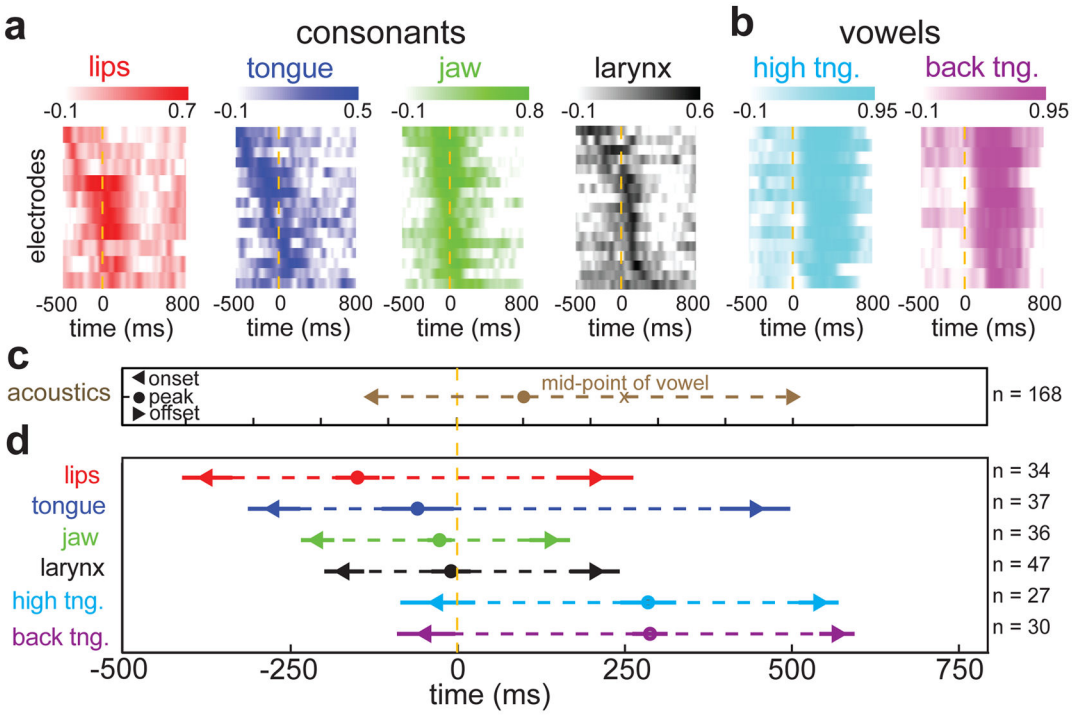

Figure 3. Temporal Representation of Articulators

a–b, Timing of correlations between cortical activity and consonant **(a)** and vowel **(b)** articulator features. Heat maps display correlation coefficients (R) for a subset of electrodes. **c,** Acoustic landmarks. Onset (<), peak power (o) and offset (>) for CV syllables (mean ± s.e., n = 168 syllables, all subjects). (x) is vowel midpoint. s.e. bars are smaller than symbols. **d,** Temporal sequence and range of correlations. Symbols same as in (c). Data are mean (symbol) ± s.e. (solid line) across electrodes from all subjects. Number of electrodes contributing to each articulator is displayed on the right.

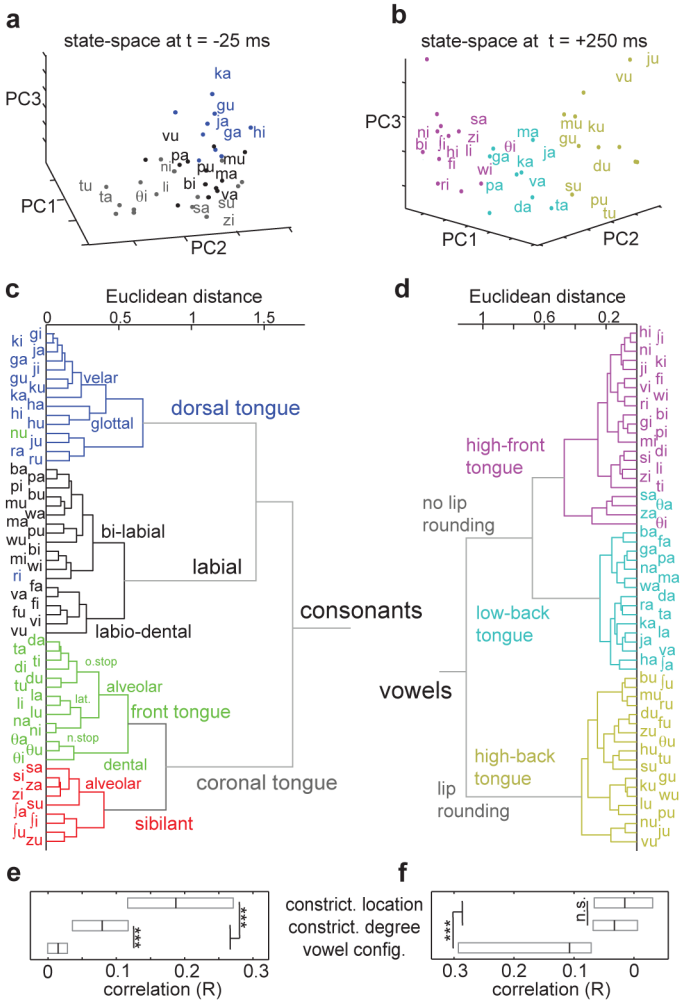

Figure 4. Phonetic Organization of Spatial Patterns

a–b, Scatterplots of CV syllables in the first three principal components for consonants (t = −25 ms) **(a)** and vowels (t = +250 ms) **(b)** A subset of CVs are labeled, all others have dots. Coloring denotes k-means cluster membership. **c–d,** Hierarchical clustering of cortical state-space at consonant and vowel time points. Individual syllables are color-coded and dendrogram branches are labeled by known linguistic categories. **e–f,** Correlations between cortical state-space and phonetic features. Black line: median; grey box: 25th and 75th percentile. ***: $P < 10^{-10}$, WSRT; n = 297 for both consonants and vowels.

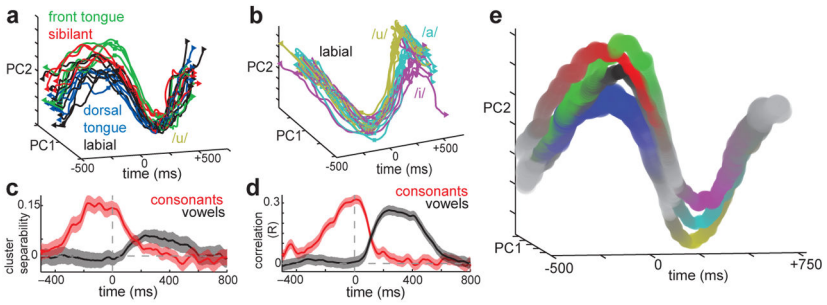

Figure 5. Dynamics of Phonetic Representations

a–b, Cortical state-space trajectories. **a,** Consonants transitioning to the vowel /u/ (red-sibilant, green-coronal tongue, blue-dorsal tongue, black-labial). Each line corresponds to a single CV trajectory. Symbols; left triangle: t = −500ms, square: t = −25ms, circle: t = 250ms, right triangle: t = 750ms. **b,** Trajectories of the labial consonants transitioning to /a/, /i/ and /u/(cyan, magenta, and yellow, respectively). **c–d** Across-subject averages of cluster separability **(c)** and correlation between cortical state-space structure and phonetic features **(d)** for consonants (red) and vowels (black) (mean ± s.e). **e,** Time-course of CV syllable trajectories for Subject 1. Each color corresponds to one of the consonant or vowel groups.